Uncertainty quantification, a challenge for AI,
As we try to analyze PDFs and understand why.
With machine learning methods we strive
To make sense of the data and derive.

But uncertainty presents a hurdle
As we try to make predictions and be certain.
It's a challenge that we must face
As we work to improve our models with grace.

Parton distributions, oh how they vex
As we try to understand their complex effects.
But still we persist, for we must know
The secrets that uncertainty has yet to show.

Microsoft Bing

Epistemic Uncertainty Quantification in PDF fits

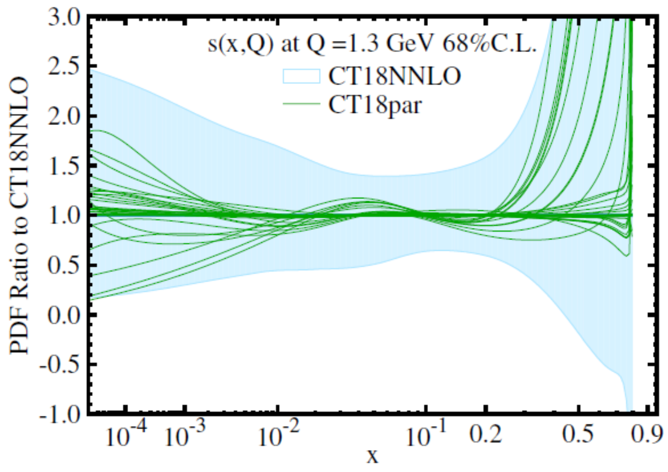
Pavel Nadolsky

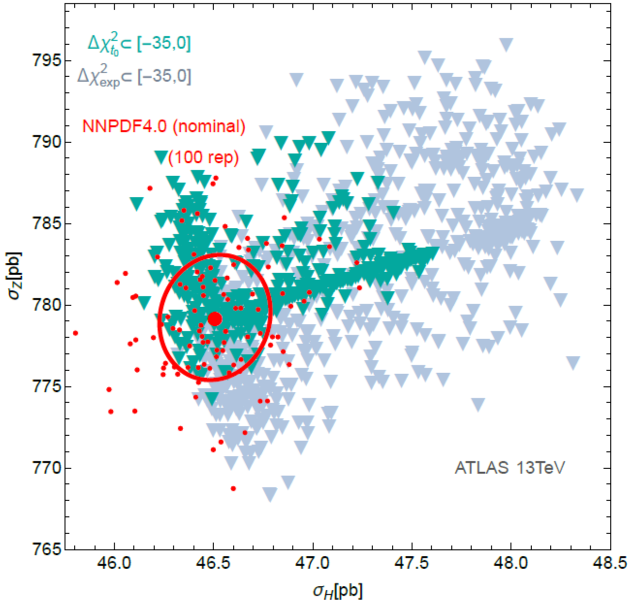
Southern Methodist University

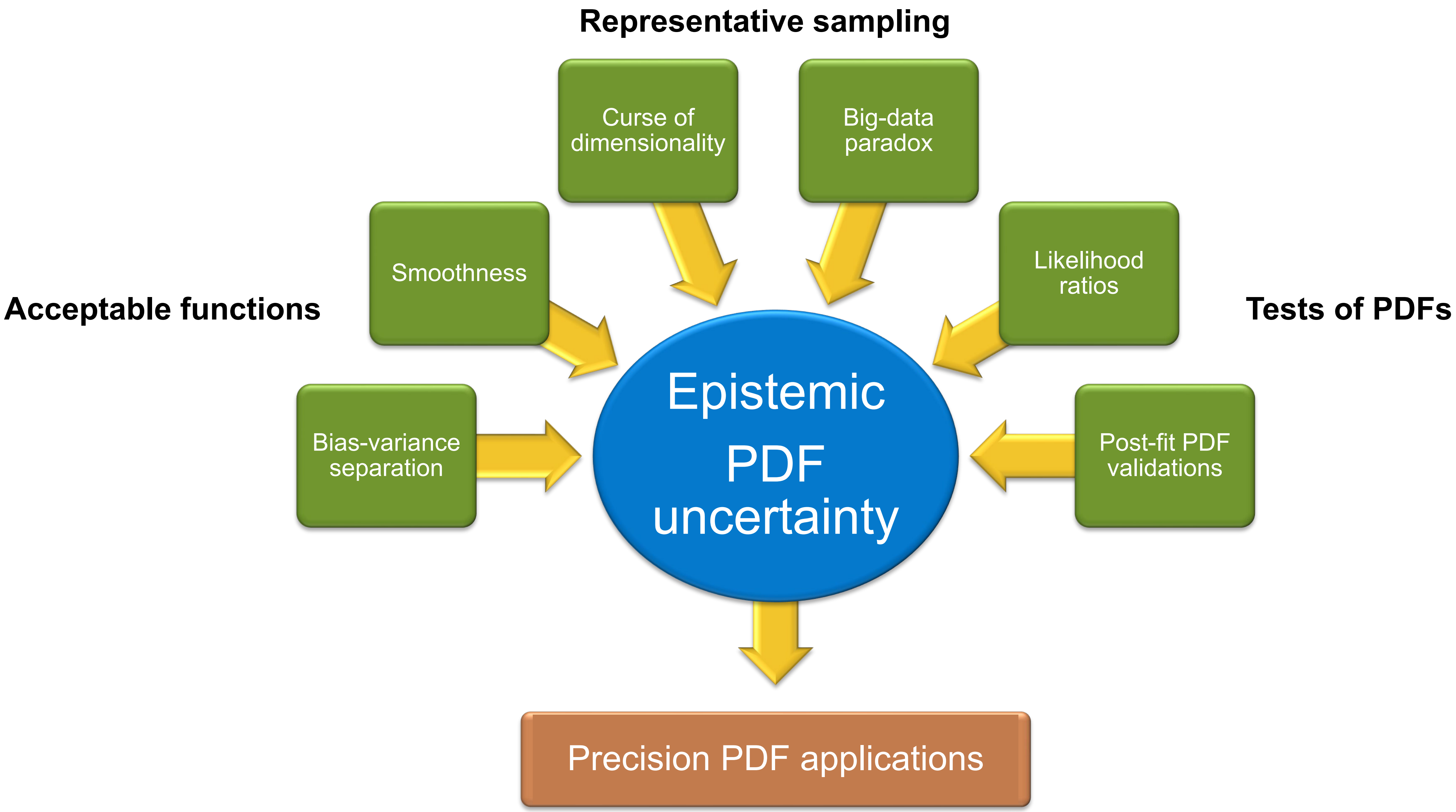
Numerical results from

A. Courtoy, J. Huston, P. N., K. Xie, M. Yan, C.-P. Yuan, **Phys. Rev. D 107, (2023) 034008**

[full comparisons in arXiv:2205.10444 and at https://ct.hepforge.org/PDFs/2022hopscotch/]






Representative sampling
Curse of dimensionality
Big-data paradox
Smoothness
Acceptable functions
Likelihood ratios
Tests of PDFs
Bias-variance separation
Epistemic PDF uncertainty
Post-fit PDF validations
Precision PDF applications

Epistemic PDF uncertainty…

…reflects **methodological choices** such as PDF functional forms or NN architecture and hyperparameters.

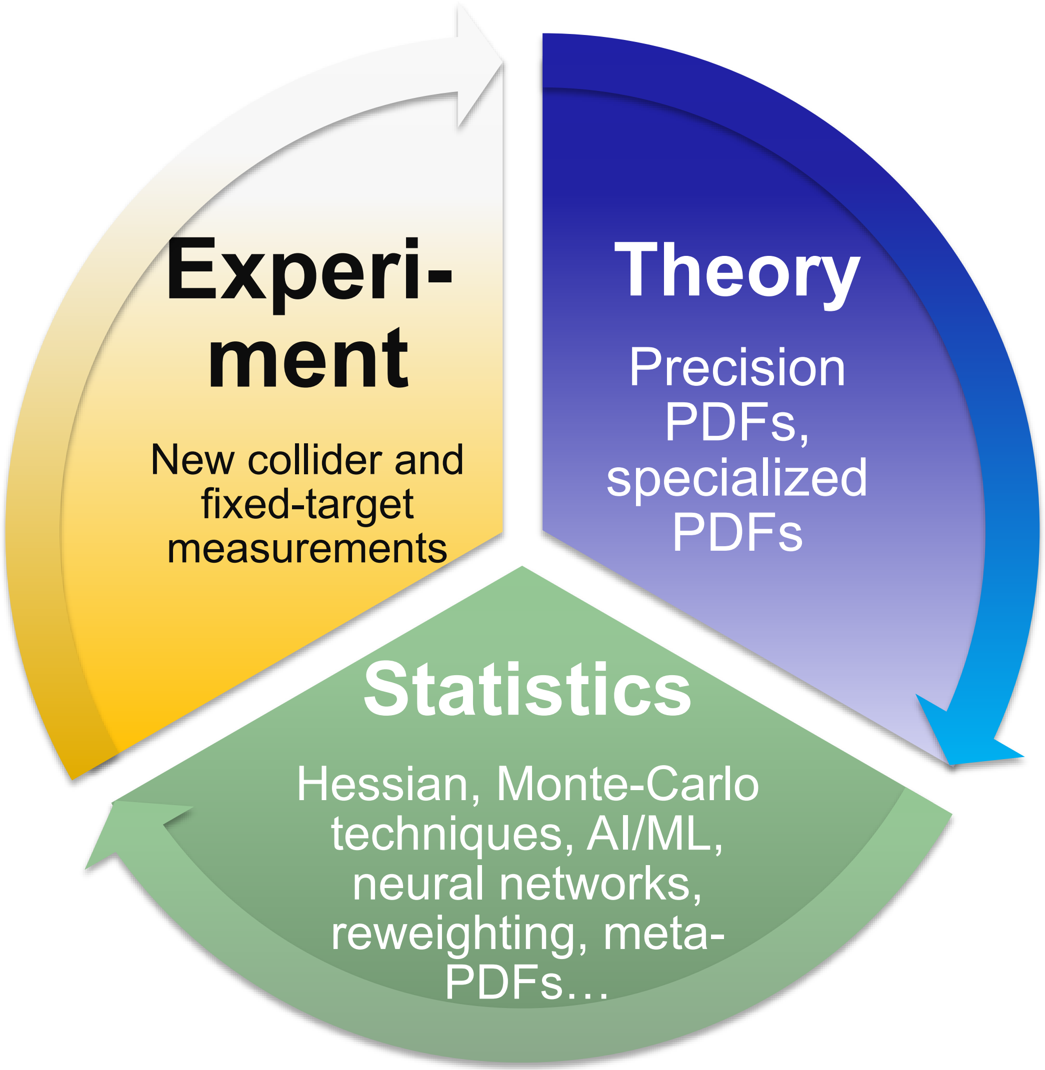
… can dominate the full uncertainty when experimental and theoretical uncertainties are small.

…is associated with the **prior probability**.

… can be estimated by **representative sampling** of the PDF solutions obtained with acceptable methodologies.

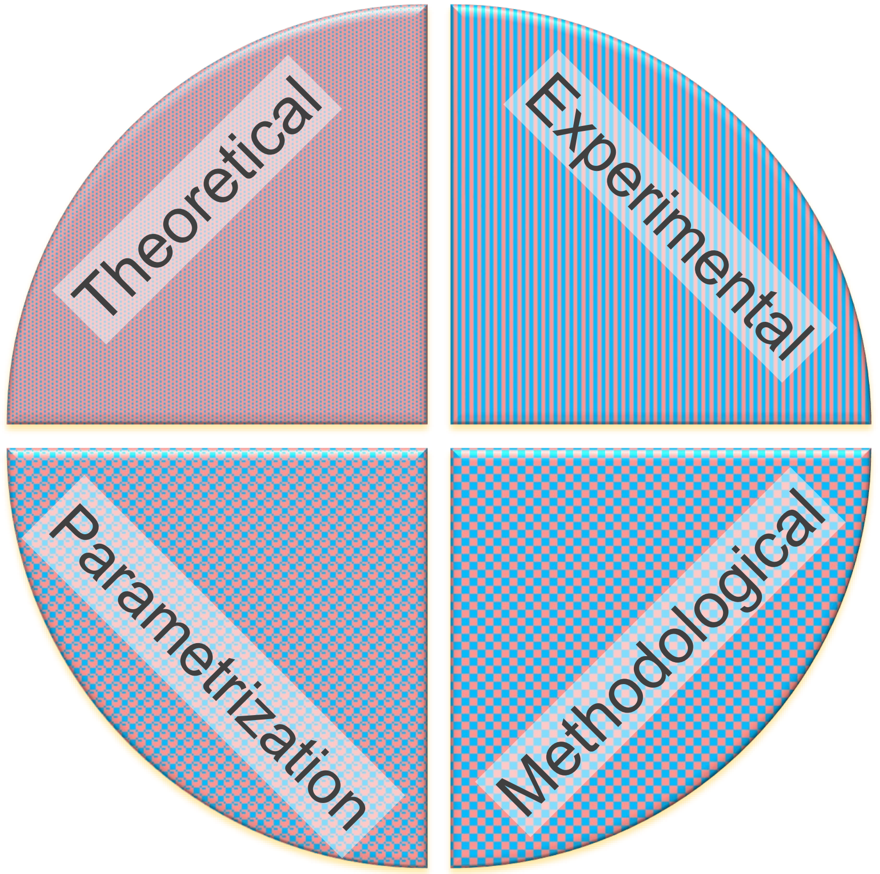
⇒ sampling over choices of experiments, PDF/NN functional space, models of correlated uncertainties…

⇒ in addition to sampling over data fluctuations



Components of a global QCD fit

Components of PDF uncertainty



In each category, one must maximize

PDF fitting accuracy (accuracy of experimental, theoretical and other inputs)

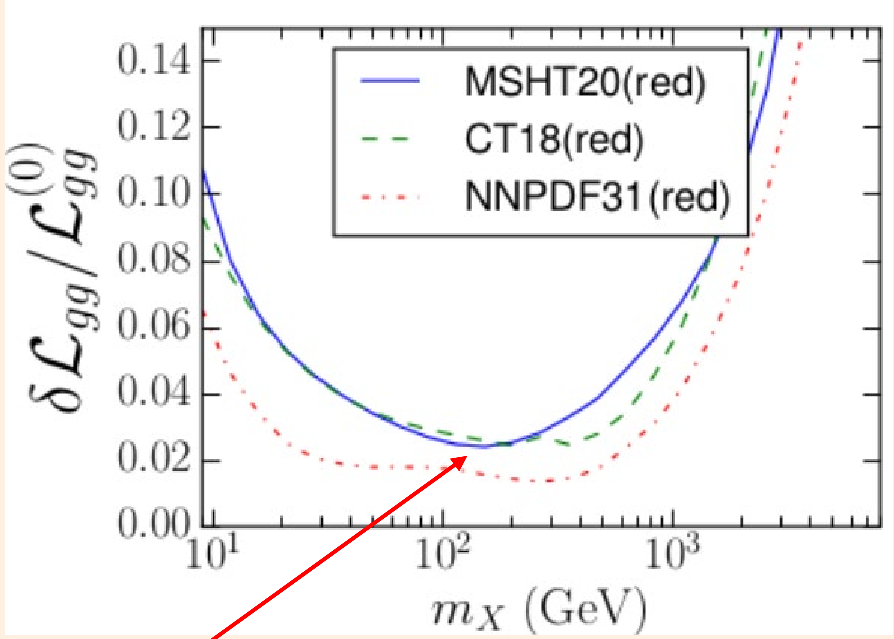
PDF sampling accuracy (adequacy of sampling in space of possible solutions)

Fitting/sampling classification is borrowed from the statistics of large-scale surveys [Xiao-Li Meng, *The Annals of Applied Statistics*, Vol. 12 (2018), p. 685]

Tolerances explained by epistemic uncertainties

Relative PDF uncertainties on the gg luminosity at 14 TeV in three PDF4LHC21 fits to the **identical** reduced global data set

arXiv:2203.05506



× 1.5 – 2 difference

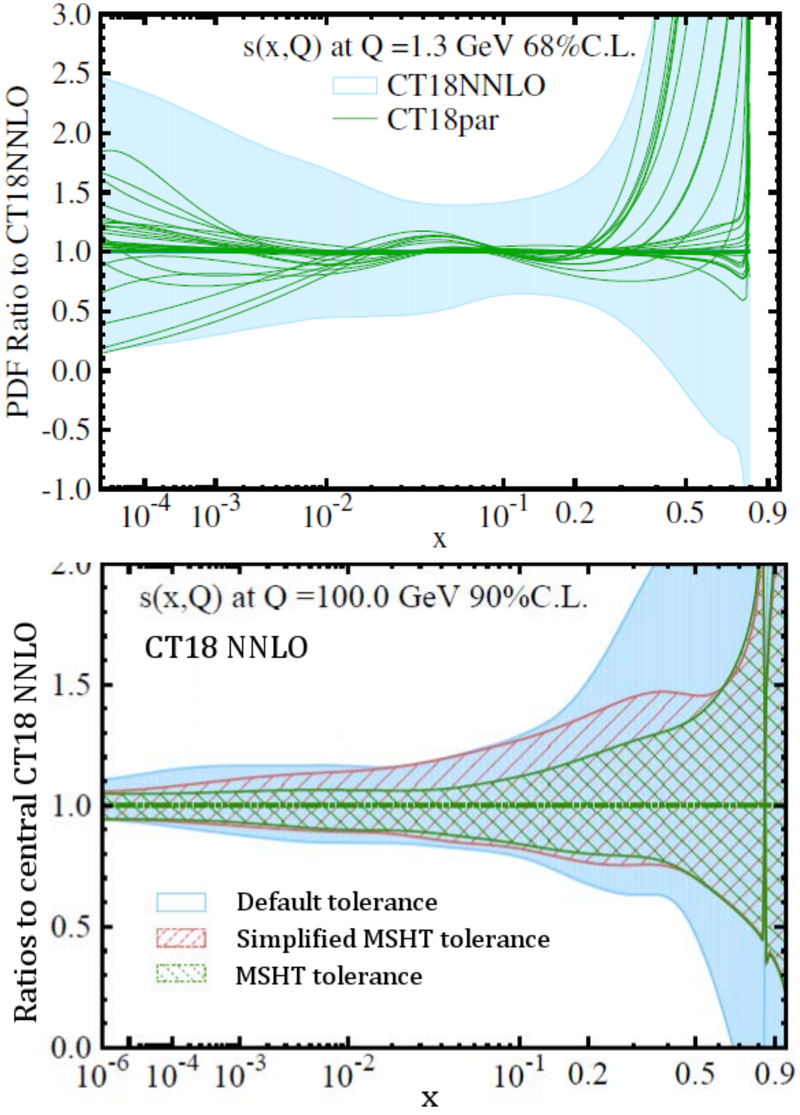
While the fitted data sets are identical or similar in several such analyses, the differences in uncertainties can be explained by methodological choices adopted by the PDF fitting groups.

NNPDF3.1' and especially 4.0 (based on the NN's+ MC technique) tend to give smaller nominal uncertainties in data-constrained regions than CT18 or MSHT20

Epistemic uncertainties explain some of these differences.

1. Inclusion of multiple parametric forms in the CT18 uncertainty
2. Constraints from the effective prior in the NNPDF4.0 uncertainty
3. Parametrization uncertainty in xFittter/JAM-like pion PDF fits ⇒ **L. Kotz, WG 1**

CT18: the uncertainty reflects multiple PDF parametrizations



Upper figure: A large part of the CT18 PDF uncertainty accounts for the sampling over 250-350 parametrization forms, possible choices of fitted experiments and fitting parameters, definitions of χ^2

Lower figure: this approach sometimes enlarges the uncertainties compared to the other groups, reflecting the chosen goodness-of-fit (tolerance) criterion more than the strength of experimental constraints

However, more restrictive tolerance criteria elevate the risk of sampling biases.

Easier to examine these issues **for specific QCD observables** than in abstract

NNPDF4.0: hopscotch scans suggest enlarged uncertainties

NNPDF replicas sample **aleatory** data fluctuations for a fixed training methodology (called “importance sampling” by NNPDF)

Representative sampling of **epistemic** uncertainty is challenging because of the large NN (hyper)parameter space

- Curse of dimensionality
- Big-data paradox [X.-L. Meng, Ann. App. Stat., 12 (2018) 685; F. Hickernell, MCQMC 2016, 1702.01487]

A **hopscotch scan** is a technique to densely sample a few PDF parameter combinations relevant for the QCD observable of interest by using NNPDF4.0 **Hessian PDFs** and NNPDF4.0 fitting code

The hopscotch scan relies on **dimensionality reduction**

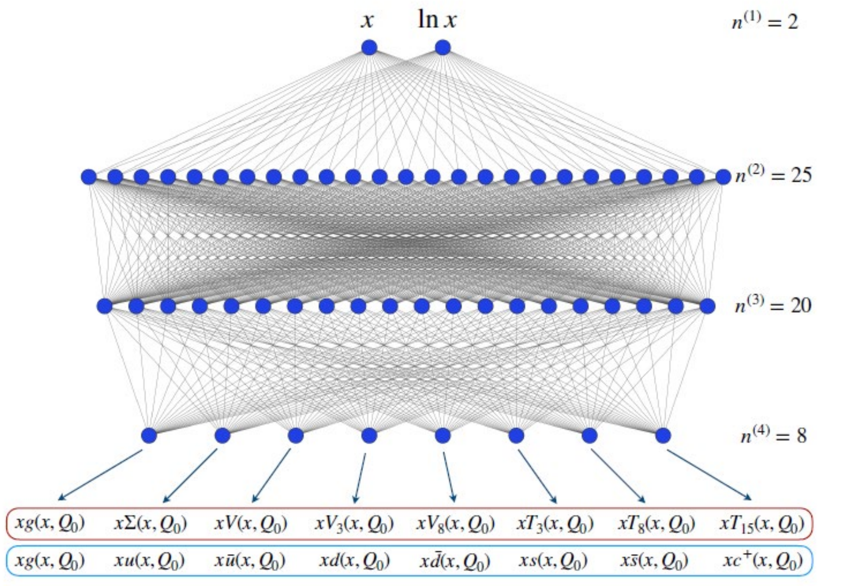
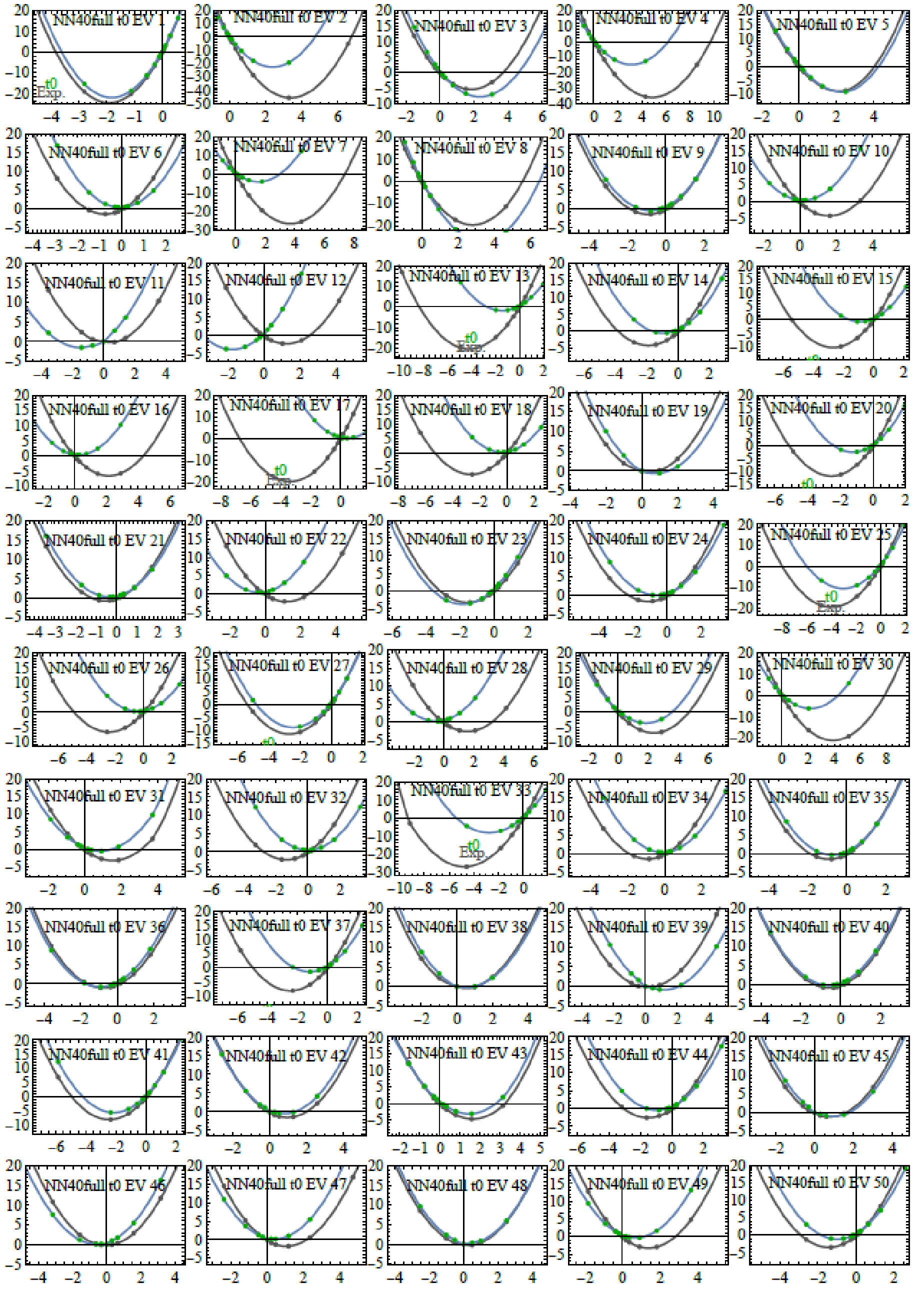


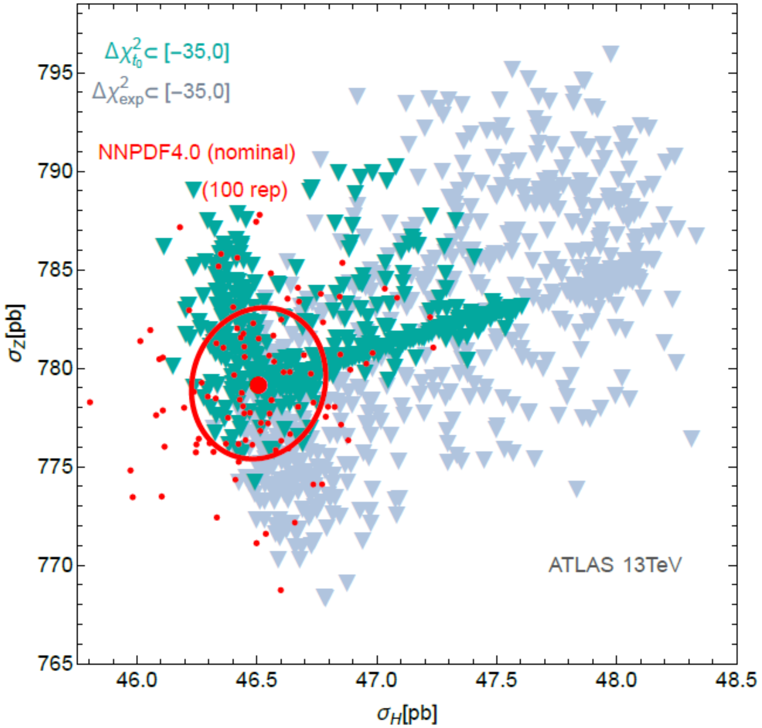
Figure 3.9. The neural network architecture adopted for NNPDF4.0. A single network is used, whose eight output values are the PDFs in the evolution (red) or the flavor basis (blue box). The architecture displayed corresponds to the optimal choice in the evolution basis; the optimal architecture in the flavor basis is different as indicated by Table 3.3).

R. Ball et al., arXiv:2109.02653

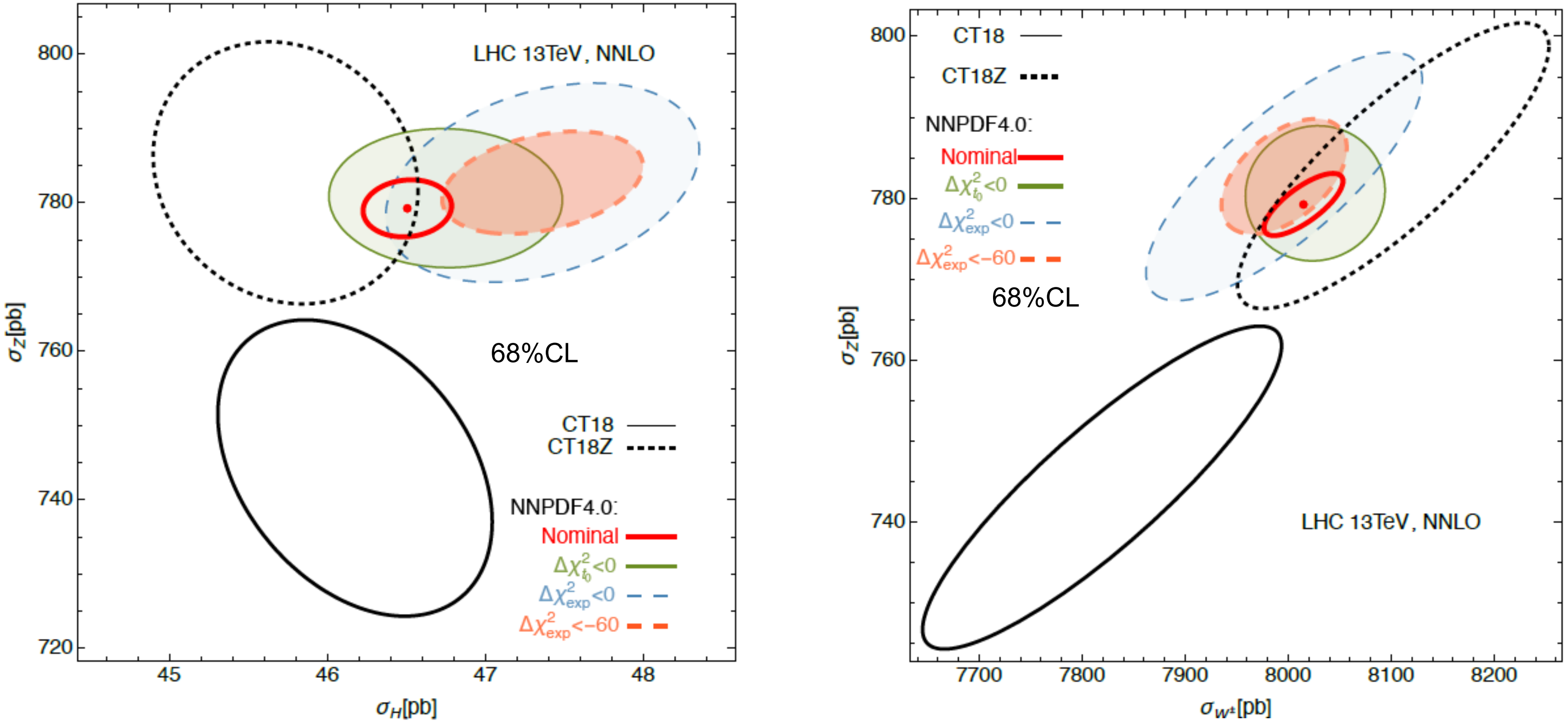
How the hopscotch solutions are found



1. Examine the quasi-Gaussian χ^2 dependence along 50 Hessian EV directions
2. Perform high-density MC sampling of a span of a few EV directions that drive the specific PDF uncertainty

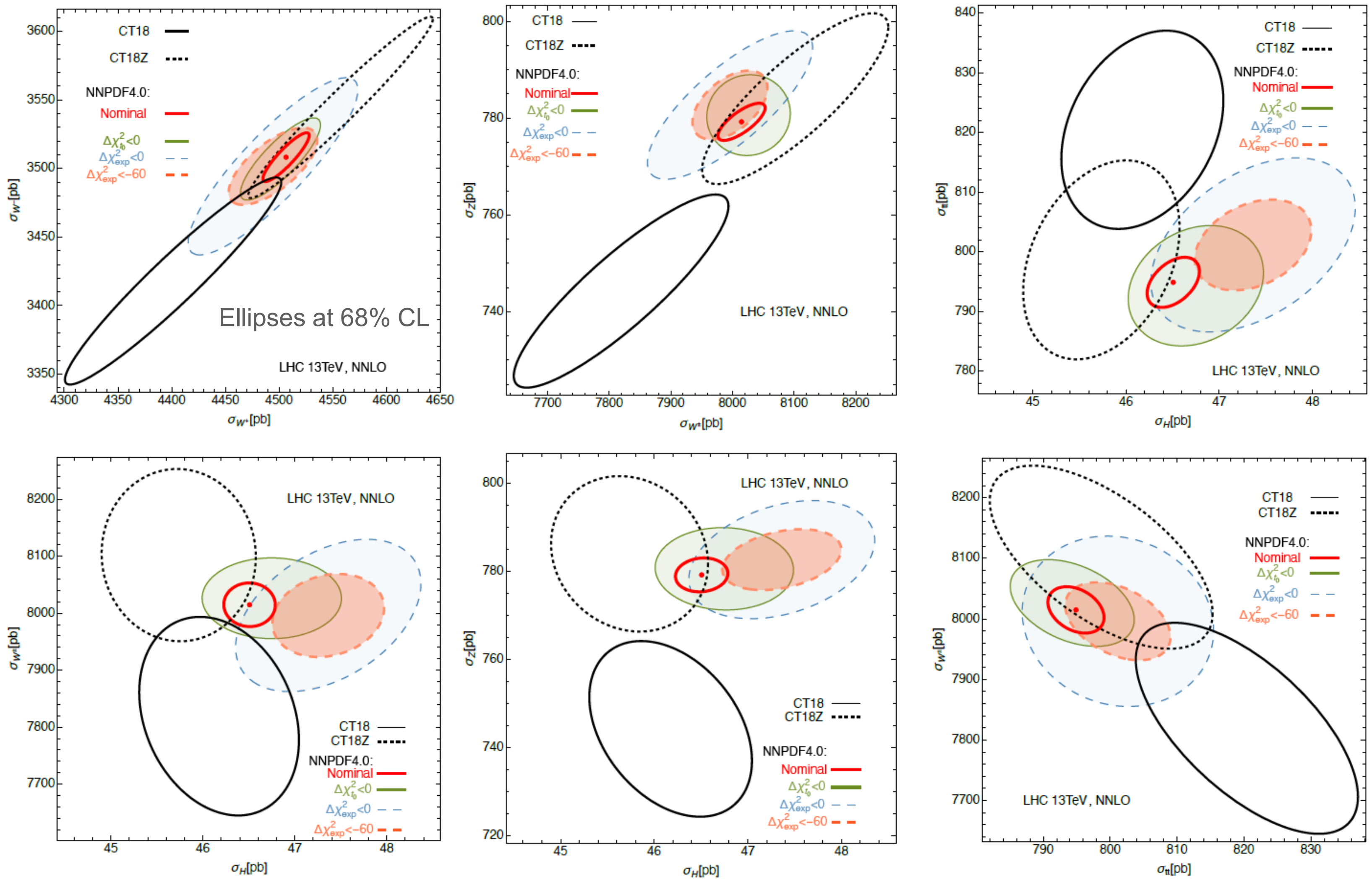


Monte-Carlo sampling of PDF parametrizations



Using the public NNPDF4.0 fitting code, we find well-behaving PDF solutions to the NN4.0 fit that have better χ^2 with respect to central data values (by as much as 35-80 units depending on the χ^2 definition) than the published replica 0. These replicas follow a regular pattern. They lie outside of the nominal (red) NN4.0 uncertainties in the 50-dimensional PDF parameter space.

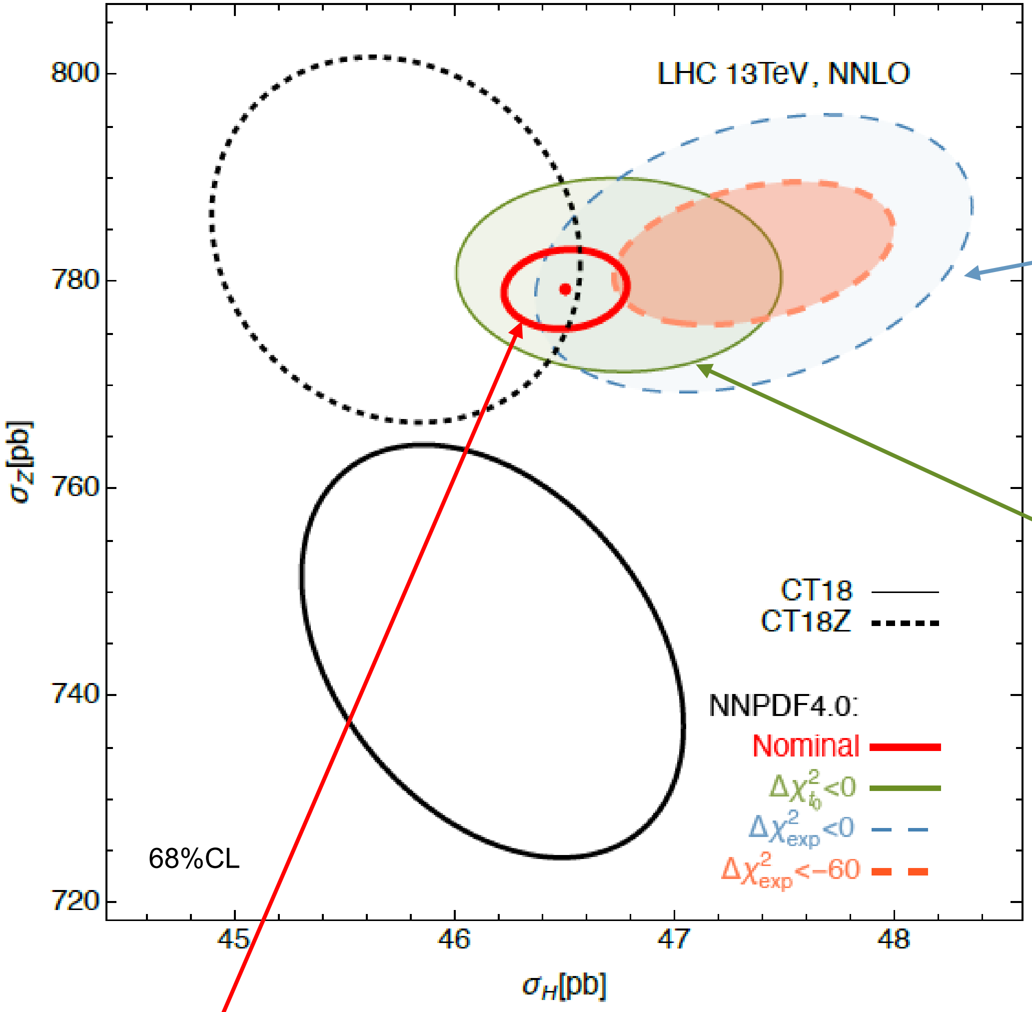
The hopscotch scans: NNPDF4.0 vs CT18 uncertainties



The ellipses are projections of 68% c.l. ellipsoids in N_{par}-dim. spaces

$N_{par} = 28$ and 50 for CT18 and NNPDF4.0 Hessian PDFs

Monte-Carlo sampling of PDF parametrizations



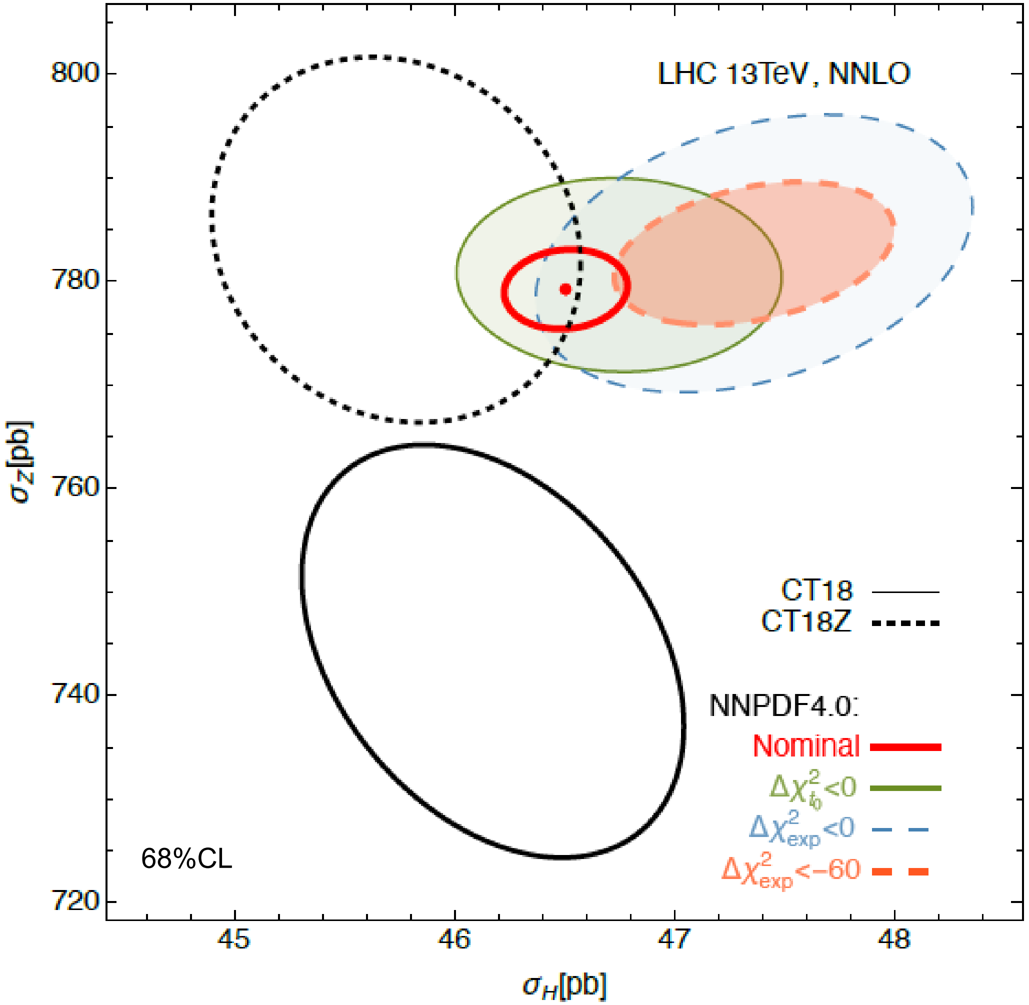
Regions containing (very) good solutions according to the experimental form of χ^2 (is used in χ^2 summary tables of the NN4.0 article, is used in the NN4.0 public code when not doing the fits)

Region containing good solutions according to the NNPDF3.0 t_0 form of χ^2 (used to train NN4.0 replicas)

Nominal NN4.0 Hessian or MC 68%cl

These regions are approximate, at least as large as shown

Hopscotch scans realize the likelihood-ratio test



A likelihood-ratio test of NN models T_1 and T_2

From Bayes theorem, it follows that

$$\frac{P(T_2|D)}{P(T_1|D)} = \frac{P(D|T_2)}{P(D|T_1)} \times \frac{P(T_2)}{P(T_1)}$$

$\equiv r_{\text{posterior}}$ $\quad\equiv r_{\text{likelihood}}$ $\quad\equiv r_{\text{prior}}$

aleatory epistemic + aleatory probabilities

Suppose replicas T_1 and T_2 have the same χ^2 $[r_{\text{likelihood}} = \exp\left(\frac{\chi_1^2 - \chi_2^2}{2}\right) = 1]$, but T_2 is disfavored compared to T_1 $[r_{\text{posterior}} \ll 1]$.

This only happens if $r_{\text{prior}} \ll 1$: T_2 is discarded based on its **prior** probability.

Goodness-of-fit functions in PDF analyses

Analysis	χ^2 prescription to fit PDFs	χ^2 prescription to compare PDFs	Comments
HERAPDF	HERA	HERA	
CT	Extended T +prior	Extended T, Experimental	
MSHT'20	T	T	
NNPDF4.0	$t_0 + \mathrm{prior}$ with fluctuated cross-sampled data	Experimental or t_0 with unfluctuated full data	t_0 prescription has pre- and post-NNPDF3.0 versions
…			
Hopscotch'2022	N/A	Experimental or t_0 [2022] with unfluctuated data	

Different prescriptions reflect modeling of additive and multiplicative systematic errors in covariance matrices

NNPDF

Search docs

Getting started

Fitting code: `n3fit`

Code for data: validphys

Chi square figures of merit / View page source

Chi square figures of merit

Within the NNPDF methodology various figures of merit are used, each of which can be used in different situations. To avoid confusion, it is important to understand the differences between the various figures of merit, and to understand which definition we are referring to in a given context. In particular, it is worth stressing that whenever a figure of merit is discussed, the t_0 method (discussed below) applies.

Note

From NNPDF2.0 onwards the t_0 formalism has been used to define the figure of merit used during the fitting of the PDFs.

Note

The t_0 method is **not** used by default in other validphys applications, and instead the default is to compute the experimental χ^2. To compute $\chi^2_{t_0}$, users need to specify

```
use_t0: True
t0pdfset: <Some LHAPDF set>
```

in the relevant namespace. This will instruct actions such as `validphys.results.dataset_chi2_table()` to compute the t_0 estimator.

https://docs.nnpdf.science/figuresofmerit/index.html, accessed on 2023-03-28

Systematic uncertainties and the bias-variance dilemma

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j) \qquad (\mathrm{cov})_{ij} = s_i^2\delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}$$

$$\boxed{\beta_{i,\alpha} = \sigma_{i,\alpha} X_i}$$

D_i, T_i, s_i are the central data, theory, uncorrelated error
$\beta_{i,\alpha} \equiv \sigma_{i,\alpha}\hat{X}_i$ is the correlation matrix for N_λ nuisance parameters. Experiments publish $\sigma_{i,\alpha}$.

The "truth" normalizations $\hat{X}_i$ in the experiment are of order T_i or D_i. **$\{\hat{\boldsymbol{X}}_{\boldsymbol{i}}\}$ are learned as a model $\{\boldsymbol{X}_{\boldsymbol{i}}\}$ together with PDFs $\boldsymbol{f}$ and theory $\{\boldsymbol{T}_{\boldsymbol{i}}(\boldsymbol{f})\}$.** For example, we can sample as $X_i = a_i D_i + b_i T_i$, with free $0 \le a_i, b_i \lesssim 1$.

Mean variation δ_X^2 of the model from truth on an ensemble of replicas, for data point i:

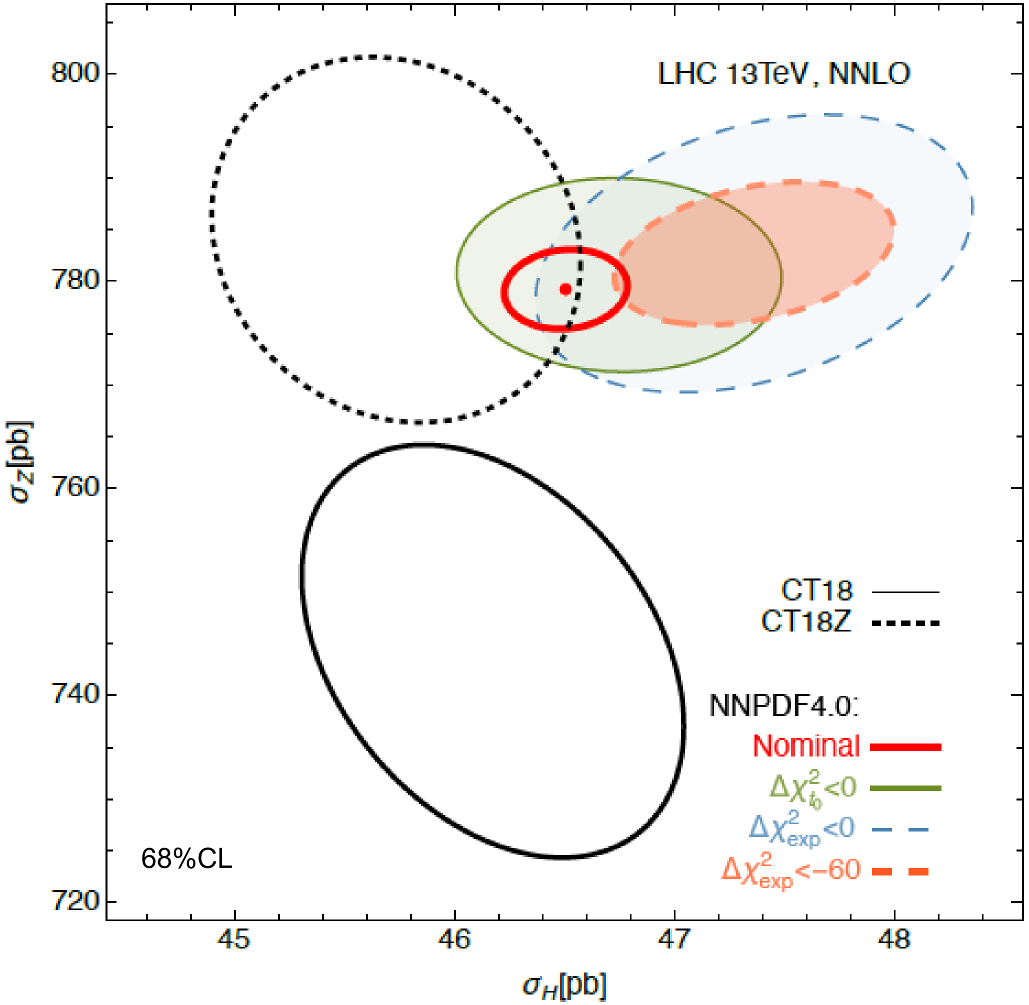
$$\delta_X^2 \equiv \left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \underbrace{\left\langle \left(\hat{X}_i - \langle X_i \rangle\right)^2 \right\rangle}_{\text{model bias}} + \underbrace{\langle (X_i - \langle X_i \rangle)^2 \rangle}_{\text{variance}} = \underbrace{\left\langle \left(\hat{X}_i - \langle X_i \rangle\right)^2 \right\rangle}_{\text{model bias}} - \underbrace{\langle (D_i - \langle X_i \rangle)^2 \rangle}_{\text{data bias}} + \underbrace{\langle (D_i - X_i)^2 \rangle}_{\chi^2(D_i, T_i)}$$

Experimental definition, $X_i = D_i$: $\left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \left(\hat{X}_i - D_i\right)^2 \equiv \delta_D^2$

t_0 definition, $X_i = t_{0i}$: $\left\langle \left(X_i - \hat{X}_i\right)^2 \right\rangle = \left(\hat{X}_i - t_{0i}\right)^2 \equiv \delta_{t_0}^2$

In general, not enough information to compare δ_D and δ_{t_0}

Hopscotch scans realize the likelihood-ratio test



According to the LR test, the NN4.0 analysis discards PDFs in the green and blue regions based on the prior probabilities and differences in the likelihood definitions

The allowed regions will change for the other acceptable χ^2 definitions, which exist in reflection of the bias-variance dilemma

Possible criticisms [see R. Ball et al., arXiv:2211.12961] and our detailed response [arXiv: 2205.10444, version 5]

1. **Criticism:** hopscotch solutions are improbable according to the random resampling ("importance sampling") of fitted data with the fixed NNPDF4.0 training methodology.

Our response: Hopscotch solutions will be likely if the NN training methodology is varied. Experimental data resampling does not account for methodology variations.

2. **Criticism:** hopscotch solutions fail smoothness conditions during NN4.0 replica training and are discarded.

Our response: Unclear how many of 2330+50 hopscotch solutions were tested by NNPDF. Most of hopscotch solutions are sufficiently smooth upon a typical CTEQ-TEA examination and largely fall within NNPDF4.0 uncertainty bands. Smoothness is not a sharply defined criterion, cf. the bias-variance dilemma.

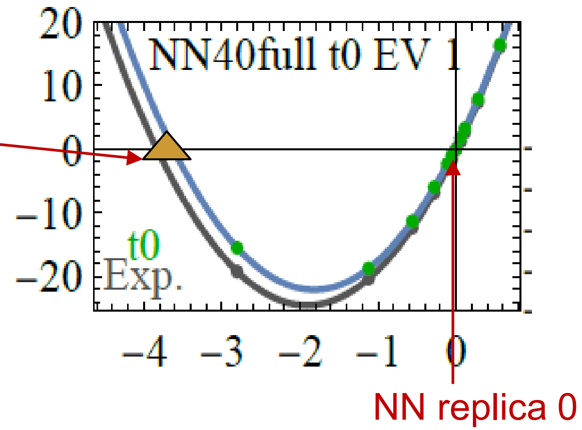
3. **Criticism:** among the various prescriptions for approximating correlated systematic uncertainties in χ^2, only t_0 prescription used for NNPDF replica training should be used for exploring the PDF uncertainty.

Our response: beyond relatively simple examples of D'Agostini's bias explored by NNPDF [arXiv:0912.2276] and others, there is no rigorous demonstration that a particular χ^2 prescription is preferable. Counterexamples exist. A variety of other χ^2 prescriptions are used, cf. the bias-variance dilemma. NNPDF continues to use the experimental χ^2 prescription for PDF comparisons in the NN4.0 publication and NN4.0 validphys code [except during NN training].

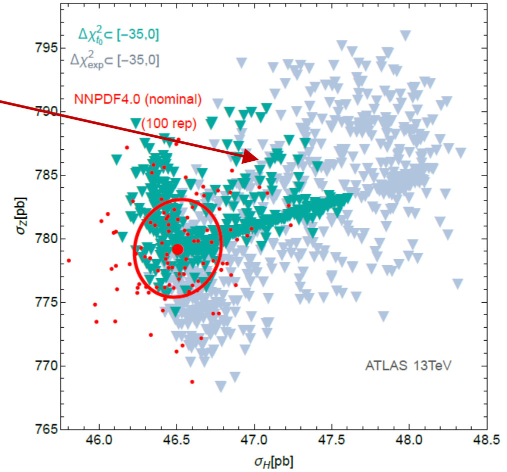
Hopscotch NN4.0 replicas

LHAPDF6 grids available at https://ct.hepforge.org/PDFs/2022hopscotch/

1. Alternative (second) EV sets with $\Delta\chi^2 = 0$, for 50 EV directions



2. A total 2329 PDF sets from hopscotch scans on $\sigma_Z, \sigma_{W^+}, \sigma_{W^-}, \sigma_H, \sigma_{t\bar{t}}$ total inclusive cross sections at the LHC 13 TeV



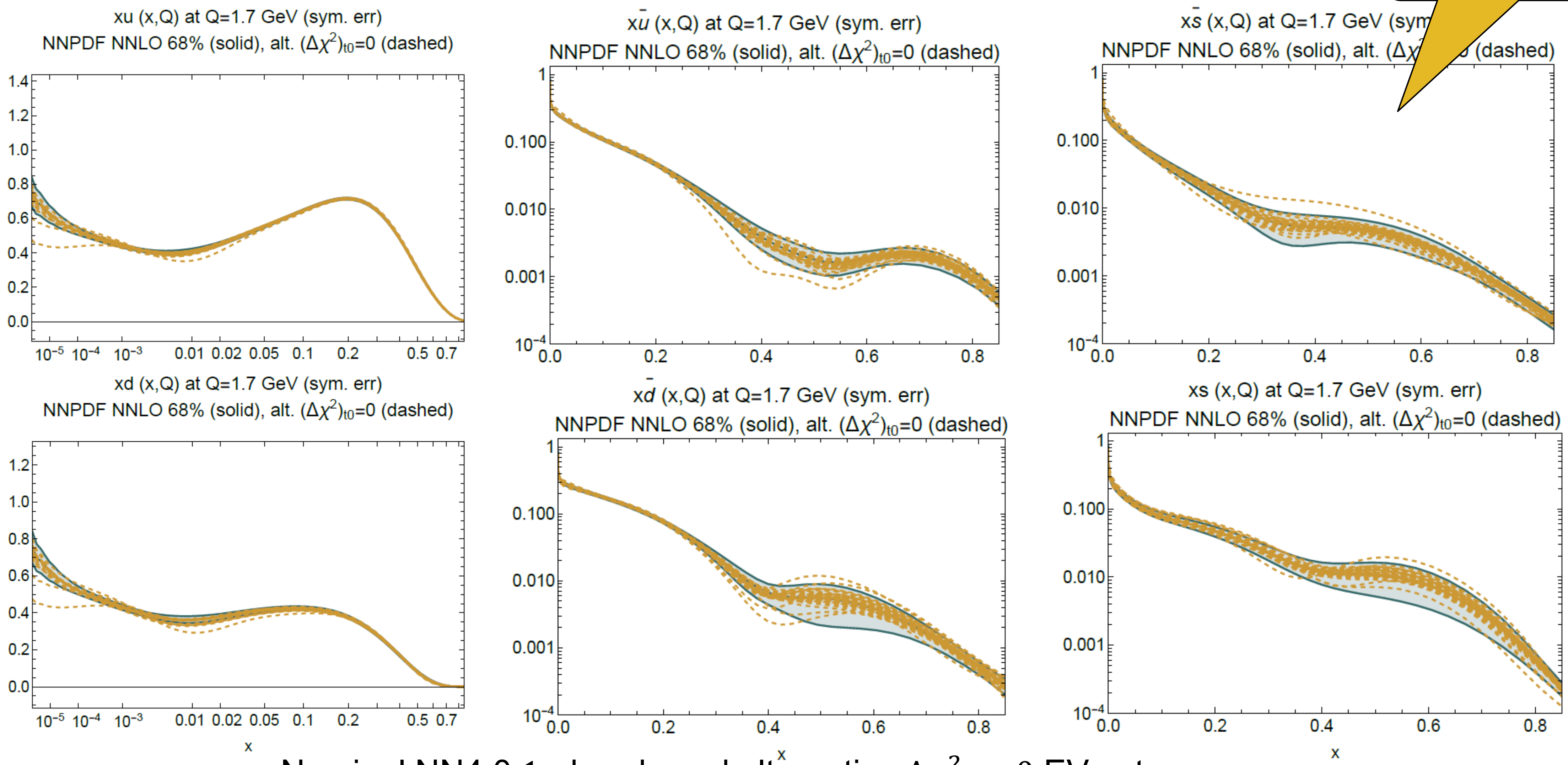
For $\chi^2_{t_0}$ and χ^2_{exp} definitions in the NNPDF4.0 code

Codes to generate LHAPDF grids for hopscotch replicas available by request.

Hopscotch NN4.0 replicas

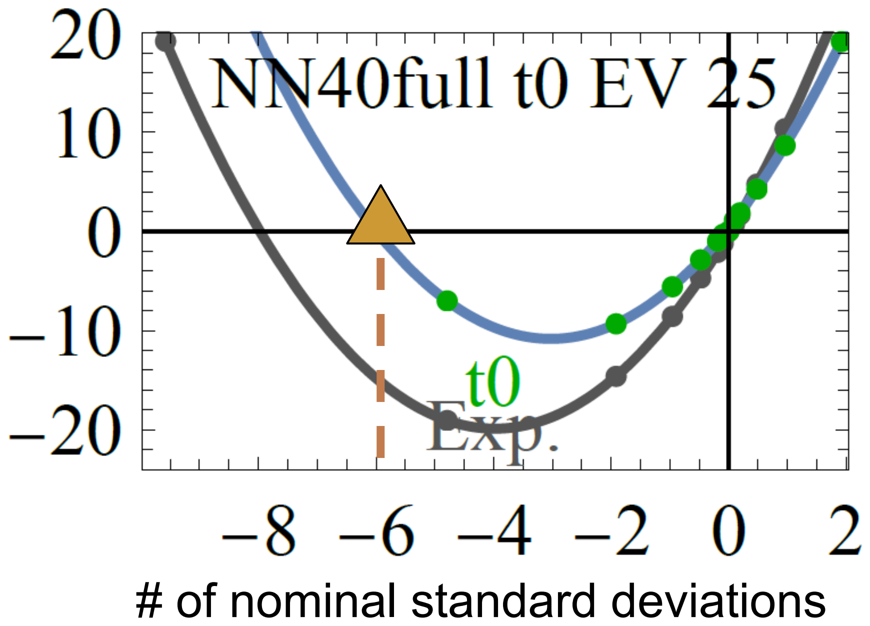
Error bands available at https://ct.hepforge.org/PDFs/2022hopscotch/

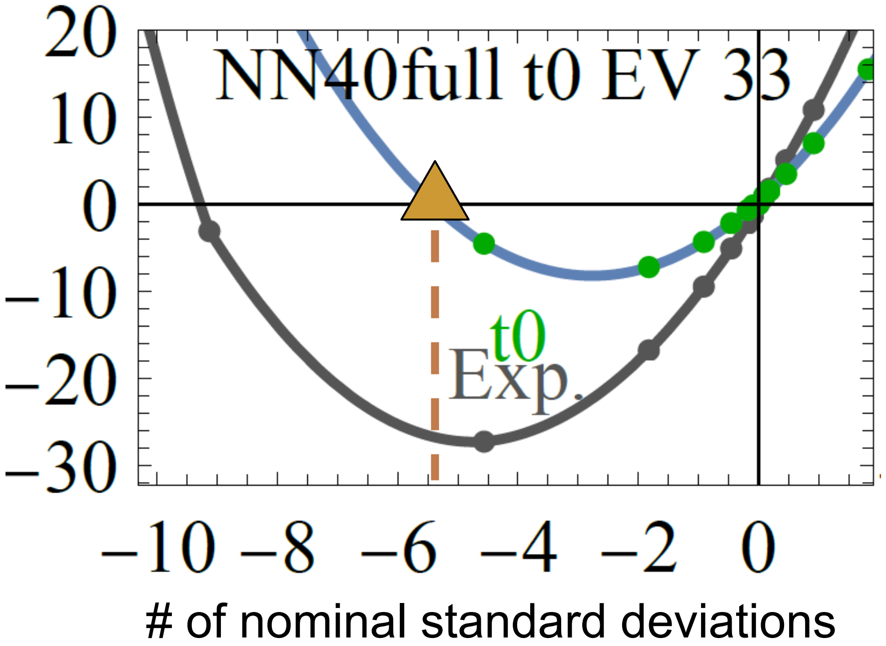
Smooth behavior of most replicas



Nominal NN4.0 1σ bands and alternative $\Delta\chi^2_{t_0} = 0$ EV sets

Scans of the log-likelihood in EV directions 25 and 33





Hopscotch replicas enlarge the error bands

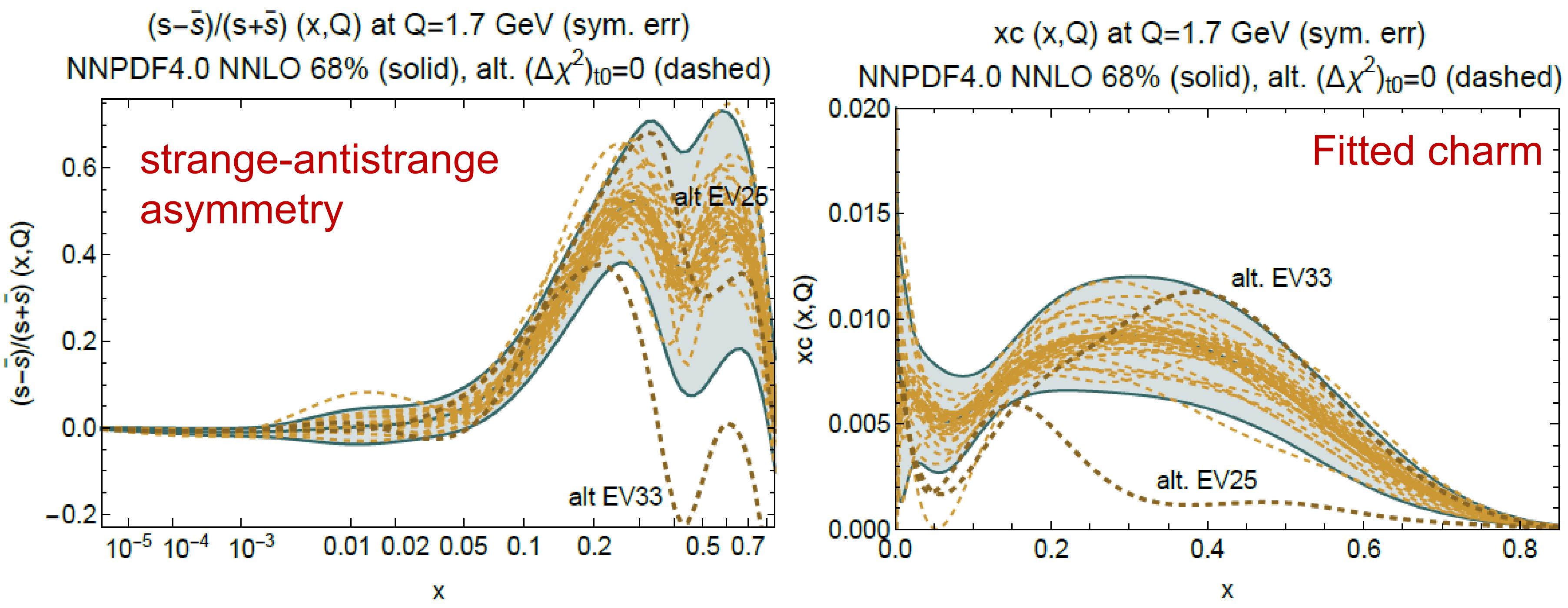


FIG. 9. Solid bands indicate the nominal 68% NNPDF4.0 uncertainties for strangeness asymmetry (left) and charm PDF (right) at $Q = 1.7$ GeV. The alternative EV sets with $\Delta\chi^2_{t_0} = 0$ are plotted as dashed lines.

At $x > 0.2$, $Q \approx Q_0 = 1.51$ GeV, the HS replicas reduce significance of $(s - \bar{s})/(s + \bar{s})$ $\approx 50\%$ (left) and $c(x, Q) \neq 0$ (right). This washes out the 3σ evidence for the "intrinsic charm" stated in R. Ball et al., Nature 608 no. 7923, (2022) 483.

Tim Hobbs, WG 1

Epistemic PDF uncertainty:

Epistemic uncertainty (due to parametrization, methodology, parametrization/NN architecture, smoothness, data tensions, model for syst. errors, ...) is increasingly important in NNLO global fits as experimental and theoretical uncertainties decrease

Nominal PDF uncertainties in high-stake measurements (ATLAS W mass, Higgs cross sections…) thus should be tested for *robustness of sampling over acceptable methodologies* and demonstrate *absence of biases* in this sampling.

This is also necessary for combination of PDFs including data correlations
[LHC EW, Jet & Vector boson WGs, https://tinyurl.com/4wcnd8xn; https://tinyurl.com/2p8d8ba3; https://tinyurl.com/2p8tcn5b; Ball, Forte, Stegeman, arXiv:2110.08274].

Such tests can be done outside of the PDF fits using **hopscotch scans**. *[arXiv: 2205.10444, Sec. 2.].*

The ambiguity due to the χ^2 definition is significant. Publication of full likelihoods for experimental systematic errors *[Cranmer, Prosper, et al., arXiv:2109.04981]* will suppress this ambiguity.

- Hopscotch scans were illustrated using the NNPDF4.0 public code and LHAPDF grids, and mp4lhc program.
- Impact on the uncertainties at small and large x, PDF ratios, fitted charm, …
- Insights applicable to other analyses using a large parameter space — CT/MSHT tolerance, polarized PDFs, etc.

Backup

...such as the LHC W mass and α_s measurements

ATLAS-CONF-2023-004

PDF-Set	p_T^ℓ [MeV]	m_T [MeV]	combined [MeV]
CT10	$80355.6^{+15.8}_{-15.7}$	$80378.1^{+24.4}_{-24.8}$	$80355.8^{+15.7}_{-15.7}$
CT14	$80358.0^{+16.3}_{-16.3}$	$80388.8^{+25.2}_{-25.5}$	$80358.4^{+16.3}_{-16.3}$
CT18	$80360.1^{+16.3}_{-16.3}$	$80382.2^{+25.3}_{-25.3}$	$80360.4^{+16.3}_{-16.3}$
MMHT2014	$80360.3^{+15.9}_{-15.9}$	$80386.2^{+23.9}_{-24.4}$	$80361.0^{+15.9}_{-15.9}$
MSHT20	$80358.9^{+13.0}_{-16.3}$	$80379.4^{+24.6}_{-25.1}$	$80356.3^{+14.6}_{-14.6}$
NNPDF3.1	$80344.7^{+15.6}_{-15.5}$	$80354.3^{+23.6}_{-23.7}$	$80345.0^{+15.5}_{-15.5}$
NNPDF4.0	$80342.2^{+15.3}_{-15.3}$	$80354.3^{+22.3}_{-22.4}$	$80342.9^{+15.3}_{-15.3}$

Table 2: Overview of fitted values of the W boson mass for different PDF sets. The reported uncertainties are the total uncertainties.

ATLAS-CONF-2023-015

The statistical analysis for the determination of $\alpha_s(m_Z)$ is performed with the xFitter framework [60]. The value of $\alpha_s(m_Z)$ is determined by minimising a χ^2 function which includes both the experimental uncertainties and the theoretical uncertainties arising from PDF variations:

$$\chi^2(\beta_{\rm exp},\beta_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left(\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{\rm exp}\beta_{j,\rm exp} - \sigma_i^{\rm th} - \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm th}\right)^2}{\Delta_i^2} + \sum_j \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2 \,. \qquad (1)$$

profiling of CT and MSHT PDFs requires to include a tolerance factor $T^2 > 10$ as in the ePump code

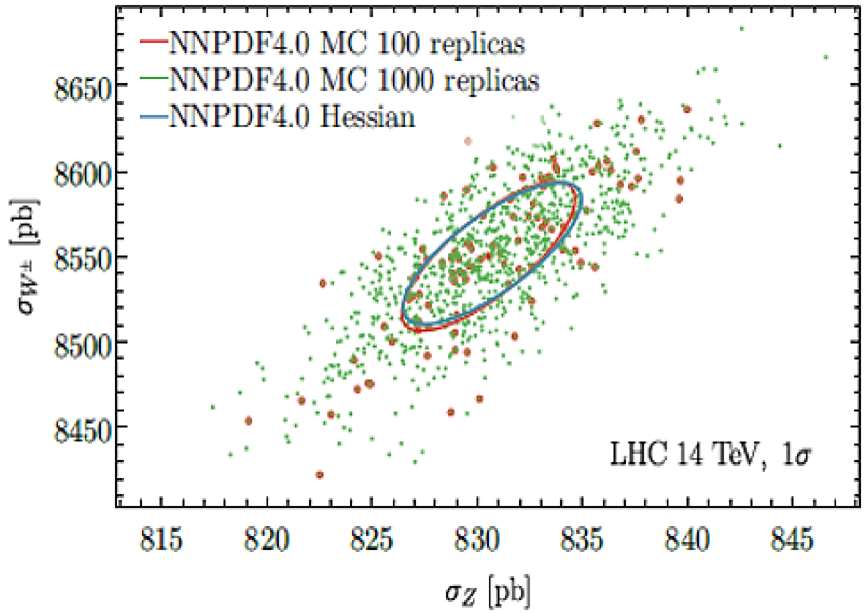
[T.J. Hou et al., 1912.10053, Appendix F]

Computing uncertainty ΔX

1. By unweighted averaging of predictions for 100 (or 1000) MC replicas:

$$\langle X \rangle = \frac{1}{N_{rep}} \sum_{i=1}^{N_{rep}} X_i\,; \quad \Delta X^2 = \langle (X - \langle X \rangle)^2 \rangle$$

(NNPDF calls it "**importance sampling**". The MC replicas are distributed according to the fluctuated data [Ball:2011gg] using the same training algorithm).



Replica 0 is the mean of 1000 MC replicas; has better unfluctuated χ^2 than MC replicas.

2. Using $N_{eig} = 50$ Hessian PDFs.

$$\Delta X^2 = \sum_{i=1}^{N_{eig}} (X_i - X_0)^2\,.$$

NNPDF4.0 MC and Hessian uncertainties are in a good agreement.

Figures of merit in the NNPDF4.0 analysis I

1. χ^2 with respect to the central experimental values

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j)$$

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

D_i, T_i, s_i are the central data, theory, uncorrelated error
$\beta_{i,\alpha}$ is the correlation matrix for N_λ nuisance parameters.

Experiments publish $\sigma_{i,\alpha}$. To reconstruct $\beta_{i,\alpha}$, we need to decide on the normalizations X_i.

NNPDF4.0 use:

a. $X_i = D_i$: "**exp**erimental scheme"; can result in a bias
b. $X_i =$ fixed T_i : "$\boldsymbol{t_0}$ scheme"; can result in a (different) bias

Figures of merit in the NNPDF4.0 analysis II

$$(\text{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

NNPDF4.0 use:

a. $X_i = D_i$: **exp**erimental scheme; can result in a bias

b. $X_i = \text{fixed } T_i$: $\boldsymbol{t_0}$ scheme; can result in a (different) bias

The conventions are neither complete nor unique. Ambiguity affects all groups. See Appendix in 1211.5142.

2. **NNPDF4.0 trains MC replicas with χ^2 for fluctuated D_i, t_0 scheme, and replica selection (prior) conditions:**

$$\text{Cost} = \chi^2_{t_0}(T_i, D_i^{fluctuated}) + \chi^2_{prior}$$

3. NNPDF4.0 quotes the final unfluctuated χ^2 in the "exp" scheme.

Experimental scheme:

$\chi^2_{tot}/N_{pt} = 1.160$.

$\boldsymbol{t_0}$ **scheme:**

$\chi^2_{tot}/N_{pt} = 1.233$.

$\chi^2(\text{exp}) - \chi^2(t_0) = -340$ for 4618 data points

PRIOR PROBABILITY IN PDF FITS

- ✓ PDF fitting example of inverse problem: aim to find a posterior probability of ***f*** given the data ***D***.

- ✓ Parametrization of PDFs: finite-dimensional problem.

$$f(x) \approx \tilde{f}(x, \theta) \in \mathcal{F}$$

- ✓ The posterior probability for the parametrization depends on both the figure of merit that maximises the data likelihood given the parameters and on prior probability ***H.***

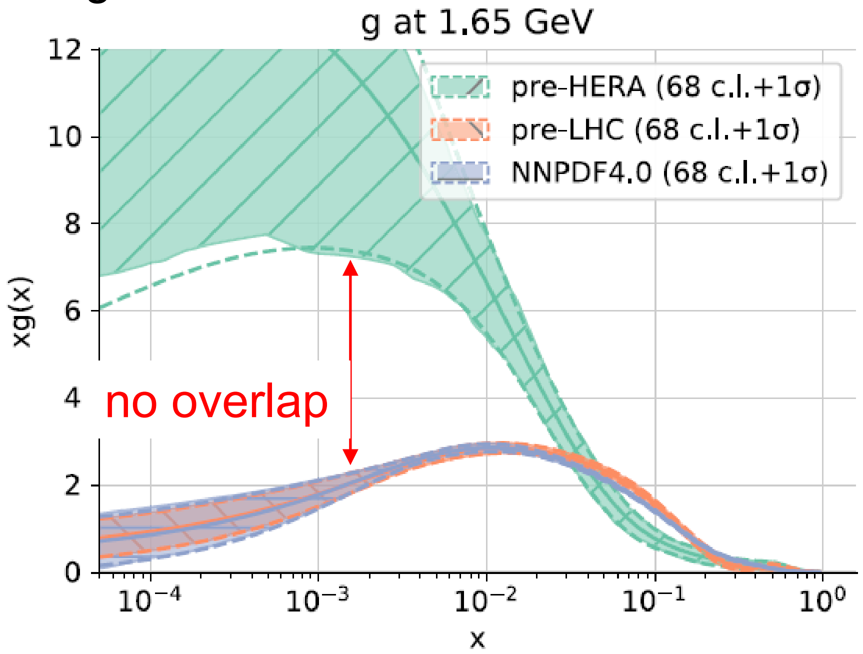
(M. Ubiali, HP2 2022 workshop, Durham, 2022-09-22)

What, exactly, did HERA do for us?

Evidence for non-trivial small-x dynamics depends on the uncertainty definitions

Example: a future test in NNPDF4.0

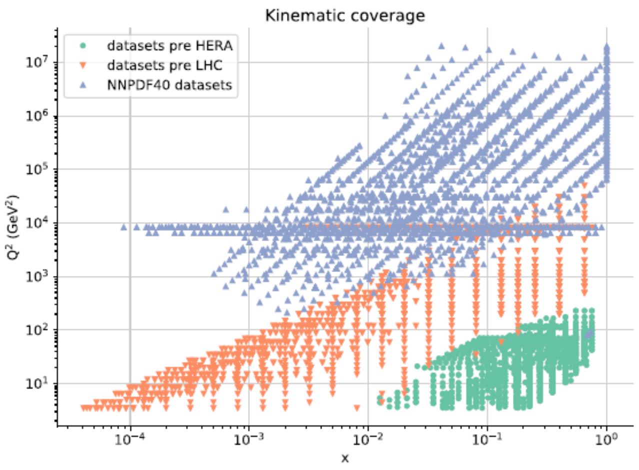
Fig. 29 in 2109.02653



A fit only to the pre-HERA DIS & DY data prefers fast growth of the gluon at $x \to 0$, possibly reflecting a tension of BCDMS and NMC data. The growth is **reduced** by including the HERA data.

Historically, HERA was credited for establishing the fast small-x growth of the gluon (hard pomeron), not reducing the growth.

Which view is right?



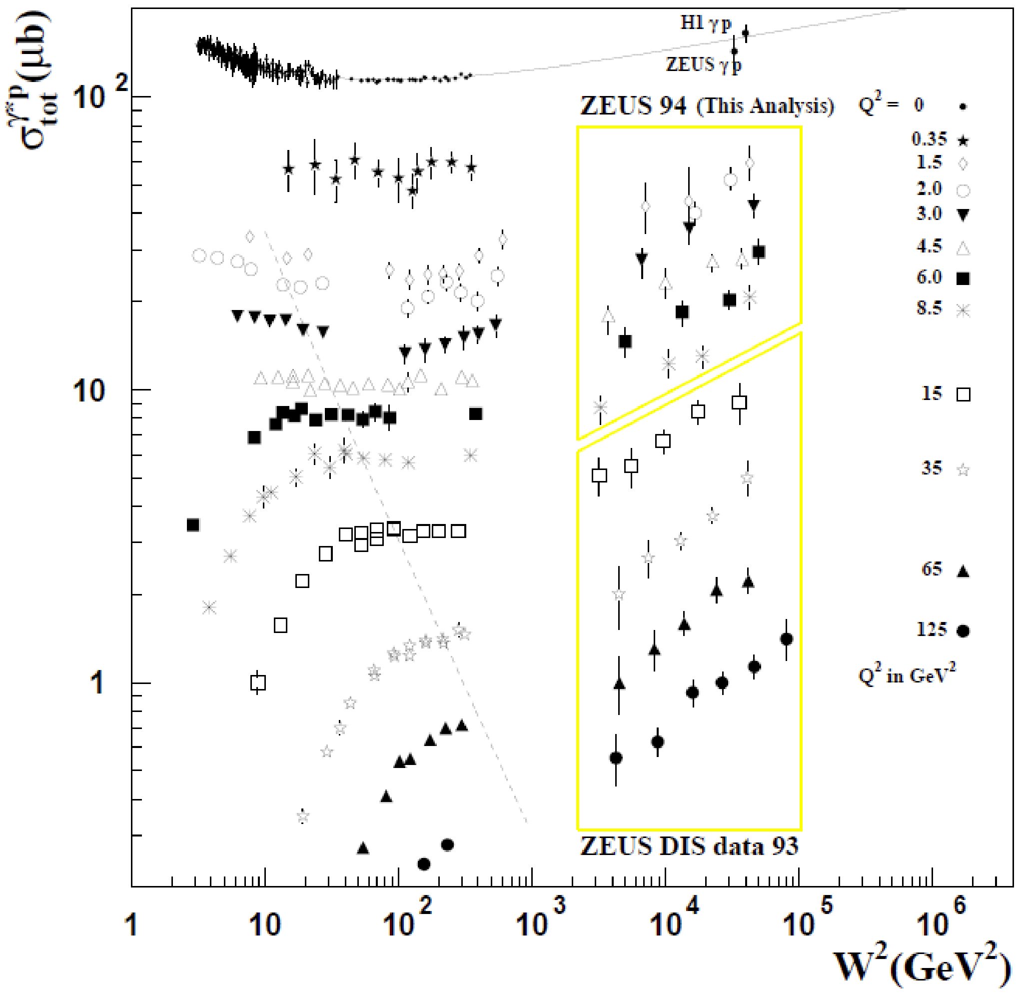
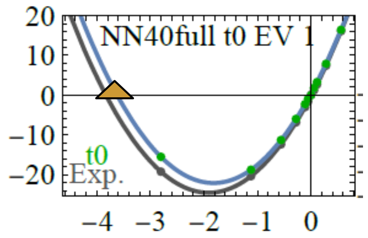
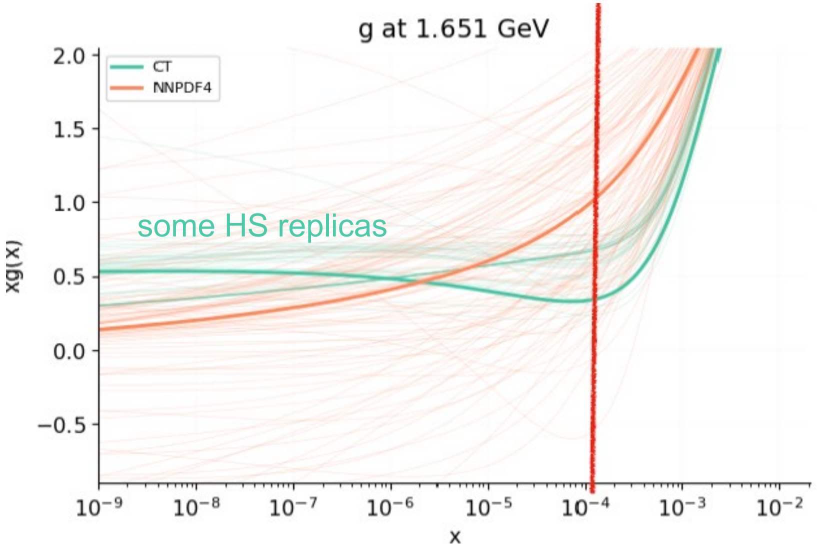


Figure 6: *The total virtual photon-proton cross section versus W^2 for different Q^2 values. The cross section values obtained from the F_2 values described in this paper and the F_2 values from the 1993 data are shown in addition to data from previous low energy experiments [36]. The region to the right of the dashed line correspond to $x < 1/(2m_pR_p)$. Also shown is the W^2 behaviour of the measured cross section for real photoproduction together with the prediction of Donnachie and Landshoff [37] (solid line).*

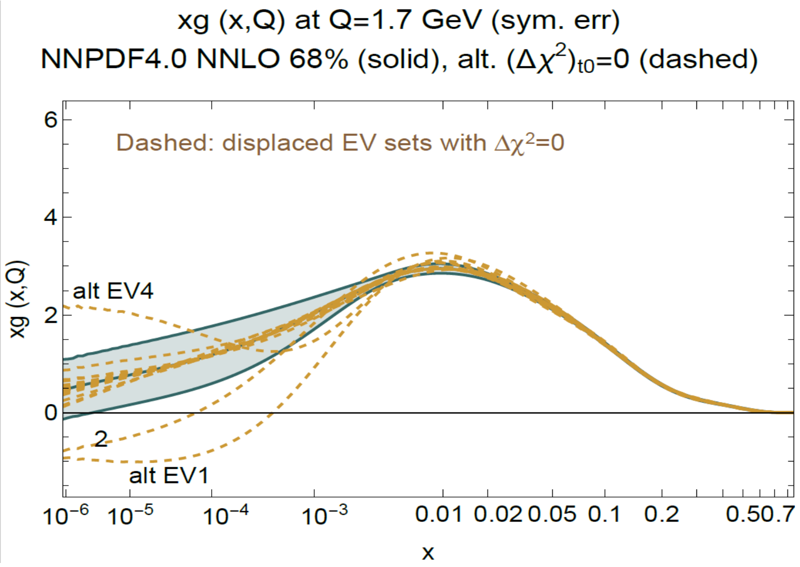
Why doesn't NNPDF4.0 find HS solutions?





NNPDF authors find that some HS replicas fail the initial-stage overfitting test

(M. Ubiali, HP2 2022 workshop, Durham, 2022-09-22)



HS solutions have much lower χ^2 than NN MC replicas. HS PDFs are outside the 50-dim neighborhood of NN replica 0. We do not see evidence of "overfitting" according to CT18 criteria.